Google Data Analyst Certificate

**Case Study: How Does a Bike-Share Navigate Speedy Success?**

horizontal line

# Placeholder image

# BACKGROUND

Cyclistic is a bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic’s offerings include reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about

8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

Cyclisitc was founded in 2016, and since then, the fleet has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

There exist three pricing plans: single-ride passes,full-day passes and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Financial analysis has shown that the annual memberships model is far more profitable than the casual rider model. However, most customers choose the ride option.

The sales manager has established that maximizing the number of annual members is the optimal growth strategy; particularly given that brand awareness is already good.

**My Role:** A junior data analyst charged with the responsibility of leveraging data analytics to inform the design of a marketing strategy

**Goal**: Re-design marketing strategies aimed at converting casual riders into annual members.

**Core Business Question**: "How do annual members and casual riders use Cyclistic bikes differently?"

**Deliverables:**

* A clear statement of the business task
* A description of all data sources used
* Documentation of any cleaning or manipulation of data
* A summary of your analysis
* Supporting visualizations and key findings
* Your top three recommendations based on your analysis

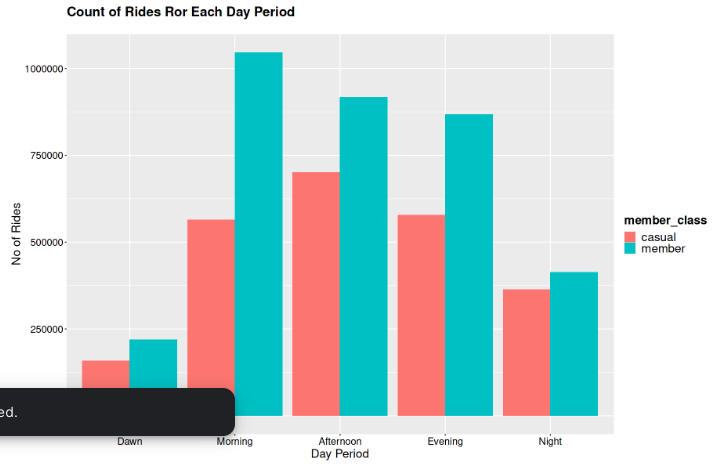
The step by step process used in the analysis is briefly described below:

## Process & Analyze

I uploaded the monthly datasets on kaggle. I then proceeded to read in the individual datasets into an R notebook. I cleaned up the data, combined the datasets, and then performed exploratory data analysis. I later performed some data aggregation and used the outcome to do some visualizations on kaggle. The procedure used in Kaggle is detailed below:

1. Uploading data: Downloading all the 35 datafiles(in zipm file format) from the AWS repository, and then uploading these as a new dataset on Kaggle
2. Setting up a new kaggle notebook with introductory notes in markdown
3. Installing relevant packages : tidyverse, lubridate,ggplot2,dplyr
4. Loading all relevant packages
5. Read in two sample files into the notebook as r dataframes, suing the R “read\_csv” function. The idea was to ensure consistency across the individual csv files in terms of column names and datatypes
6. Read in the files for the last 12 months into a single r dataframe using the “map\_df” function
7. Intiti
8. Created new columns for:
9. Ride Length - did this by subtracting end\_at time from start\_at time
10. Day of the Week
11. Month
12. Day
13. Year
14. Time - convert the time to HH:MM:SS format
15. Hour
16. Season - Spring, Summer, Winter or Fall
17. Time of Day - Night, Morning, Afternoon or Evening
18. Cleaned the data by:
    1. Checking for missing(null values)
    2. Checking for duplicates
19. Added new fields such as:
    1. Ride\_length /duration: subtracting end time from start-time
    2. Extracted hours, day periods, months, seasons from the start\_time field
       1. Hours: 1-24
       2. Day periods: Dawn, Morning, Afternoon,Evening,Night
       3. Months: 1-12
       4. Seasons: Winter,Spring,Summer,Fall
20. Renamed some fields to ensure ease of readability
21. Further data cleaning: Dropped rows with negative ride durations
22. Data Analysis & Visualization: Visualized the data using the ggplot library, and summarized the findings and inferences. The following aspects were visualized
    1. Count of bike types used overall
    2. Count of rides for each category
    3. Bike types used for each member category
    4. Number of weekday rides for each member category
    5. Number of rides for each member category for each day period
    6. Number of rides for each member category for each Season
    7. Median duration for each member category
    8. Median duration for each member category for each weekday
23. Trimmed off some columns and downloaded the resulting dataframe as csv to use in creating tableau visualizations.

**Sample R Visualizations**

****

**Tableau**

I then proceeded to create a simple tableau dashboard. Check out the screenshot below

****